DTIC FILE COPY

Arithmetic Skills in Using Algorithms

AD–A226 272

Sarah Lichtenstein and Donal MacGregor

Perceptronics, Inc.

for

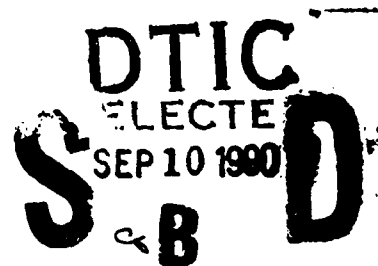**Contracting Officer's Representative
Michael Drillings**

**Basic Research
Michael Kaplan, Director**

June 1990

DTIC
ELECTE
SEP 10 1990
S B D

**United States Army
Research Institute for the Behavioral and Social Sciences**

90 09 07 011

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction
of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract for
the Department of the Army

Perceptronics, Inc.

Technical review by

Nehama Babin

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS --			
2a. SECURITY CLASSIFICATION AUTHORITY --			3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.			
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE --						
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Perceptronics PDRIR-1129-8-87			5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 90-30			
6a. NAME OF PERFORMING ORGANIZATION Perceptronics, Inc.		6b. OFFICE SYMBOL (If applicable) --	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute			
6c. ADDRESS (City, State, and ZIP Code) 6271 Variel Avenue Woodland Hills, CA 91367			7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600			
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences		8b. OFFICE SYMBOL (If applicable) --	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-83-C-0198			

8c. ADDRESS (City, State, and ZIP Code)

5001 Eisenhower Avenue
Alexandria, VA 22333-5600

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61102B	74F	n/a	n/a

11. TITLE (Include Security Classification)

Arithmetic Skills in Using Algorithms

12. PERSONAL AUTHOR(S)
Lichtenstein, Sarah; and MacGregor, Donald

13a. TYPE OF REPORT Interim	13b. TIME COVERED FROM 83/04 TO 87/08	14. DATE OF REPORT (Year, Month, Day) 1990, June	15. PAGE COUNT 20

16. SUPPLEMENTARY NOTATION Michael Drillings, contracting officer's representative

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Knowledge retrieval
			Decomposed quantitative judgments

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

An algorithm is a series of steps or operations that, when sequentially applied, produces a solution to a problem. Properly applied, algorithms are helpful when a complex or difficult numerical question can be broken into subquestions. This paper identifies and details a serious barrier to the effective use of algorithms: weak mathematical skills. Given an algorithm of 13 steps requiring copying, converting from percentages to proportions, adding, subtracting, multiplying, and dividing, 78 percent of our subjects made one or more errors in the task. These results suggest that decision aids be tested on the population for which they were intended, to avoid problems arising from unexpected deficits in users' knowledge and skills.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☑ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Michael Drillings	22b. TELEPHONE (Include Area Code) (202) 274-8722	22c. OFFICE SYMBOL OBR

An algorithm is a series of steps or operations that, when
sequentially applied, produces a solution to a problem. Properly
applied, algorithms are helpful when a complex or difficult numerical
question can be decomposed into sub-questions. Answers are given to
the sub-questions; these components are then recomposed, via the
algorithm, to arrive at an answer to the original, target question.

In a series of experiments, we have been exploring the techniques
of algorithmic decomposition as an aid to numerical estimation
(MacGregor, Lichtenstein, & Slovic, in press; Lichtenstein & MacGregor,
1984; Lichtenstein, MacGregor & Slovic, 1987; Lichtenstein & Weathers,
1987). Although the techniques lead to improvements, some subjects are
led seriously astray by the very methods that are intended to help
them.

We have previously reported on one source of subjects' poor
performance: misinformation. We asked subjects to estimate apparently
obscure numerical facts (such as the number of pounds of potato chips
consumed yearly in the U.S.) by decomposing each question into a series
of questions the answers to which are easier to estimate (e.g., pounds
of potato chips consumed per capita per week, number of weeks in a
year, and population of the U.S.). The success of such an approach
relies in part on the subjects' knowledge of these easier elements.
But we found substantial amounts of misinformation (Lichtenstein,
1987). For example only 33% of our subjects knew how many feet there
are in a mile; only 57% estimated the population of the United States with
an eight-number digit. Seriously erroneous beliefs (e.g., that the U.S.

1

population is three billion) can doom the effectiveness of algorithmic decomposition.

The present paper reports on another source of problems in using algorithms: weakness in arithmetic skills. In order to exclude the problem of misinformation, we focus here on an algorithm that requires no estimation skills. This algorithm is based on the use of Bayes' Theorem to solve a class of problems in combining probabilistic evidence; these problems have been called base-rate problems (see, e.g., Bar-Hillel, 1980). The two problems we used are shown in Table 1. The problems have different cover stories but are structurally the same.

Insert Table 1 about here

Subjects. The subjects were 76 paid volunteers who responded to ads in the University of Oregon student newspaper. The present task was completed along with several other unrelated paper-and-pencil tasks in a one- to two-hour period. The subjects were run in groups in a large university classroom. Each subject received one of the two problems shown in Table 1.

Instructions and Algorithm. The instructions said:

In this task we would like you to work through a problem by carefully following a number of detailed steps. First, you will read through the problem. Then, you will follow a series of steps, some that ask you to pull information directly from the problem itself, and others that ask you to

2

Table 1

The Base Rate Problems

Light Bulb

Consider the following problem:

A light bulb factory uses a scanning device which is supposed to put a mark on each defective bulb it spots in the assembly line. Eighty-five percent (85%) of the light bulbs on the line are OK; the remaining 15% are defective.

The scanning device is known to be accurate in 80% of the decisions, regardless of whether the bulb is actually OK or actually defective. That is, when a bulb is good, the scanner correctly identifies it as good 80% of the time. When a bulb is defective, the scanner correctly marks it as defective 80% of the time.

Suppose someone selects one of the light bulbs from the line at random and gives it to the scanner. The scanner marks this bulb as defective.

What is the probability that this bulb is really defective?

(table continues)

Table 1 (continued)

Dyslexia

Dyslexia is a disorder characterized by an impaired ability to read. Two percent (2%) of all first graders have dyslexia. A screening test for dyslexia has recently been devised that can be used with first graders. The screening test is cheap and easy to administer; it identifies those children who will later be given a more extensive test to determine for sure whether the child has dyslexia. The screening test is not completely accurate. For children who really have dyslexia, the screening test is positive (indicating dyslexia) 95% of the time. But it also gives a positive (dyslexia) result for 5% of the normal children, the ones who do not have dyslexia.

A first grader is given the screening test and the result is positive, indicating dyslexia.

What is the probability that the child really has dyslexia?

---

carry out basic arithmetic. Please follow all the directions carefully. Pay special attention to the accuracy of your arithmetic. This is not a test of your ability to do arithmetic, but accuracy of computation is essential to what we are asking you to do.

[The problem followed.]

After the problem was an algorithm composed of thirteen steps, as shown for the Lightbulb problem in Table 2. On the page following the algorithm, two additional questions were asked:

Do you think the answer in (M) is a sensible answer to the question, "What is the probability that this lightbulb is really defective [the child really has dyslexia]?

Yes _____ No _____

If you answered No, what do you think is a sensible answer? _____

Insert Table 2 about here

Results. The correct answer (to two-digit accuracy) for the Lightbulb problem is .41; for the Dyslexia problem, .28. Only 17 of the 76 subjects (22%) gave the correct answer. Table 3 shows the answers the subjects gave, categorized according to ranges around the correct answer:

Too Low: Answers falling more than .10 below the correct answers.

About Right: Answers within $\pm$.10 of the correct answer, including all correct answers.

Table 2

Algorithm for the Lightbulb Problem

(A) Out of 1,000 light bulbs produced by the factory, how many are

defective? Multiply the percentage of defective bulbs by

1,000. (First convert the percentage value to a decimal value

before multiplying.)

$$1,000 \times \underline{\hspace{4cm}} = \underline{\hspace{2cm}} \text{(A)}$$

Proportion of

Defective Bulbs

(B) Subtract your estimate in (A) from 1,000 to get the number of

bulbs out of 1,000 that are NOT defective.

$$1,000 - \text{(A)}\underline{\hspace{3cm}} = \underline{\hspace{2cm}} \text{(B)}$$

(C) What percentage of the time is the scanner

able to correctly identify light bulbs that

are actually defective? (from the problem) $\underline{\hspace{2cm}}$(C)

(D) What percentage of the time is the scanner

able to correctly identify light bulbs that are

actually not defective? (from the problem) $\underline{\hspace{2cm}}$(D)

(table continues)

6

Table 2 (continued)

(E)   Look over the following table:

LIGHT BULBS ARE:

	Actually Defective	Not defective
Scanner Says IS Defective	Box # 1	Box # 4
Scanner Says IS NOT Defective	Box # 2	Box # 3

_____ (L)

_____ + _____

(A)   (B)

(F)   Write the number of defective light bulbs from (A) on the line
      labeled (A) in the table above, just below Box #2.

(G)   Write the number of non-defective light bulbs from (B) on the
      line labeled (B) in the table above, just below Box #3.

(H)   Multiply the percentage value in (C) by your estimate from
      (A).   (First convert the percentage value to a decimal value
      before multiplying.)

                    (A) _____ x (C) _____ = _____(H)

Write your value for (H) in Box #1.

(table continues)

7

Table 2 (continued)

(I) Subtract your value in (H) from your value in (A).

(A)_____ - (H) _____ = _____(I)

Write your value for (I) in Box #2.

(J) Multiply the percentage value in (D) by your estimate from (B). (First convert the percentage value to a decimal value before multiplying.)

(B) _____ x (D) _____ = _____(J)

Write your value for (J) in Box #3.

(K) Subtract your value in (J) from your value in (B).

(B) _____ - (J) _____ = _____(K)

Write your value for (K) in Box #4.

(L) Add the numbers in Boxes #1 and #4.

Box #1_____ + Box #4_____ = _____(L)

Write your value for (L) on the line labeled (L), to the right of the boxes.

(M) To get the final answer, divide your value in Box #1 by your value for (L).

Box #1_____ ÷ (L) _____ = _____(M)

Too High: Answers that are more than .10 above the correct answer but below 1.00 (no subject got an answer of exactly 1.00).

Outside: Negative answers and answers greater than 1.00.

None: No numerical answer given.

Insert Table 3 about here

Is your answer sensible? As shown in Table 3, most subjects thought that their answers were sensible. Base-rate problems are notorious for having nonintuitive answers, so it is perhaps not surprising that subjects who arrived at about the right answer were less likely to think their answer was sensible (55%) than subjects who arrived at answers that were within the range but too high or too low (77%). Most discouraging is that more than half the subjects whose answers fell outside the bounds of 0 to 1 were satisfied with their answers.

Of the 26 subjects who said their answer was not sensible, only 20 gave a new answer. Only one of these revised answers was close to being correct; this subject had completed the algorithm perfectly, arriving at the answer of .41 to the light bulb problem, but said that a sensible answer was .35. Twelve of the 20 revised answers fell in the Too Low category, supporting the finding (shown in Table 3) that most (82%) of the subjects who had originally calculated a low answer found it sensible.

Errors. The subjects made numerous errors in the task. These errors, categorized and tallied, are shown in Table 4.

9

Table 3. Frequency of responses.

	Frequency of Answer	Frequency of Sensibleness		
		Yes	No	Blank
Too Low	18	14	3	1
About Right	29	16	13	-
Too High	9	6	3	-
Outside	17	10	7	-
None	3	-	-	3
	76	46	26	4

10

Insert Table 4 about here

The Dyslexia group were particularly prone to the error of taking

the wrong information from the story. The Dyslexia story differs from

the Lightbulb story by expressing the two pieces of diagnostic

information in two different ways:

> . . . For children who really have dyslexia, the screening
>
> test is positive (indicating dyslexia) 95% of the time. But
>
> it also gives a positive (dyslexia) result for 5% of the
>
> normal children, the ones who do not have dyslexia.

Many subjects were apparently confused by this wording, so that when

the algorithm asked, "What percentage of the time is the screening test

able to correctly identify children that actually do not have

dyslexia?" (emphasis in the original), 45% of the subjects filled in

5%. One subject even went so far as to write us a note in the margin

saying that this question was incorrectly worded.

The algorithm several times required subjects to copy a previous

calculation into a new spot. About a third of the subjects made errors

in following these simple directions.

We categorized arithmetic errors as (a) errors in sign, (b)

addition or subtraction, or (c) multiplication or division. Within the

multiplication or division errors we further distinguished decimal

errors, upside-down division, and other errors. Upside-down division

11

Table 4

Error Analysis, in Percentages

	All (n=76)	Lightbulb (n=29)	Dyslexia (n=47)
Wrong Info from Story	47	28	60
One or More Copying Error	32	28	34
One or More Arithmetic Error	54	34	66
Sign Error	8	7	9
Addition or Subtraction	13	14	13
Multiplication or Division	51	31	64
Decimal Error	22	17	26
Upside-down Division	13	3	19
Other	30	21	36
Incomplete Algorithm	4	3	4
No Errors	22	41	11
Mean No. Errors per Subject		2.21	2.89
Most Errors by One Subject		14	12

is the calculation of the inverse of the indicated divison, for example:

$$50 \div 2 = .04 \quad \text{or} \quad 2 \div 50 = 25 .$$

The Dyslexia subjects showed a significantly greater frequency of one or more arithmetic errors, $\chi^2 = 7.03$, $p > .01$, a finding we cannot explain.

An incomplete algorithm received only one tally for that reason, regardless of how many steps were omitted.

The errors made by the Algorithm subjects sometimes led to absurd answers; as shown in Table 3, 22% of the answers were outside the range of permissible probabilities, either negative or greater than 1.00. The largest answer was 4934.4; this subject made three decimal errors, one copying error, one multiplication error, one sign error, and ended with an upside-down division.

Additional Data. An additional group of 102 subjects were given the same problems with a different aid. Before reading the problem, these subjects read a lengthy (six single-spaced pages) tutorial designed to teach the subjects how to solve base-rate problems. The method presented was based on the 2-by-2 table that formed the center of the algorithm, but in contrast with the algorithm, the tutorial emphasized understanding and common sense (for further details, see Lichtenstein & MacGregor, 1984). After the tutorial each subject received one of the two problems with a worksheet. The Lightbulb version of this worksheet is shown in Table 5. As may be seen, it is shorter and requires the subjects to make judgments.

Of these 102 subjects, 35 received the task in the usual large

13

Table 5

Worksheet for the Tutorial Group

Please work the following problem using the method just described. We've drawn you a table to work with.

A light bulb factory uses a scanning device which is supposed to put a mark on each defective bulb it spots in the assembly line. Eighty-five percent (85%) of the light bulbs on the line are OK; the remaining 15% are defective.

The scanning device is known to be accurate in 80% of the decisions, regardless of whether the bulb is actually OK or actually defective. That is, when a bulb is good, the scanner correctly indentifies it as good 80% of the time. When a bulb is defective, the scanner correctly marks it as defective 80% of the time.

Suppose someone selects one of the light bulbs from the line at random and gives it to the scanner. The scanner marks this bulb as defective.

What is the probability that this bulb is really defective?

Step 1. Draw a table. Done.

Step 2. Label the table.

Step 3. Assign an arbitrary grand total. Use 1,000.

Step 4. Estimate the population totals. First decide which set of information is population information. Then divide the 1,000 into two parts, using information from the problem.

Step 5. Fill in the cells. Divide each of your estimated totals among its two cells, according to the information in the problem.

Step 6. Cross out the false. Cross out the two cells that are contradicted by the information given in the problem.

Step 7. Find the needed probability. Write the relevant numbers in the top and bottom of the fraction and convert the fraction to a decimal answer.

$$\frac{\text{\# in target cell}}{\text{Sum of \#'s in both cells}} = \underline{\qquad} = \underline{\qquad} = \underline{\qquad} , \text{ answer.}$$

classroom groups. The other 67 were run in small groups of 4 to 7 people, with fewer other tasks and with small, battery-operated calculators available for use.

Thirty-one percent of these subjects arrived at the right answer, the same percentage among those given the task in a large classroom and among those who were run in small groups.

The format of the worksheet for these subjects did not permit as detailed an analysis of errors. Arithmetic errors were found for 43% of the large-group subjects (not much less than the 54% arithmetic error rate for the Algorithm subjects) and 18% for the small-group subjects (who were encouraged but not required to use calculators).

Discussion

This paper has identified and detailed a serious barrier to the effective use of algorithms: weak mathematical skills. Given an algorithm of 13 steps requiring copying, converting from percentages to proportions, adding, subtracting, multiplying, and dividing, 78% of our subjects made one or more errors in the task.

One should not generalize these results to the U.S. population at large. These subjects were, with few, if any, exceptions, college students at a state university. As such, they are above average in intelligence and education. But they may be reasonably representative of many groups for whom decision aids are designed, such as business people, government employees, and military personnel. Our results, therefore, should be taken to heart by all those who design decision aids. The problems such designers should face are exacerbated by our

15

previous findings that this same population of subjects often hold erroneous knowledge of ordinary facts (Lichtenstein, 1987).

In this age of $5 electronic calculators and $500 computers, lack of arithmetic skills might seem unimportant. But some of the errors our subjects made are unlikely to be cured by the availability of such tools. An "upside-down" division (e.g., 2/50 = 25) can be performed easily on a calculator. And given the sentence "[The screening test] gives a positive (dyslexia) result for 5% of the normal children, the ones who do not have dyslexia," 45% of our subjects answered "5%," instead of "95%," to the question, "What percentage of the time is the screening able to correctly identify children that actually do not have dyslexia?" Electronic calculators will be of no help for such misunderstandings.

Great care should be taken, we conclude, that decision aids be tested on the population for which they are intended, to avoid as much as possible problems arising from unexpected deficits in users' knowledge and skills.

References

Bar-Hillel, M. (1983). The base-rate fallacy in probability judgments. *Acta Psychologica*, *44*, 211-233.

Lichtenstein, S. (1987). *Factual knowledge of Oregon college students* (Report No. 87-4). Eugene, OR: Decision Research.

Lichtenstein, S., & MacGregor, D. (1984). *Structuring as an aid to performance in base-rate problems* (Report No. 84-16). Eugene, OR: Decision Research.

Lichtenstein, S., MacGregor, D., & Slovic, P. (1987). *Creating algorithms as an aid to judgment* (Report No. 87-5). Eugene, OR: Decision Research.

Lichtenstein, S., & Weathers, A. G. (1987). *Creating algorithms as an aid to judgment: Part II* (Report No. 87-6). Eugene, OR: Decision Research.

MacGregor, D., Lichtenstein, S., & Slovic, P. (In press). Structuring knowledge retrieval: An analysis of decomposed quantitative judgments. *Organizational Behavior and Human Decision Processes*.